Poster III-19

PubMatrix: A Web-Based Automated Literature Mining Tool
Cheadle, Chris[1], Hosack, Doug[2], Lempicki, Richard A.[2], Engel, Jim[3], Bright, Tiffani J.[4], Becker, Kevin G.[4]
[1]Cellular Biochemistry Section, CCR/NCI, Bethesda, MD, USA; [2]Laboratory of Immunopathogenesis and Bioinformatics, NIAID, Frederick, MD, USA; [3]NCTS; [4]DNA Array Unit, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

Molecular experiments using multiplex strategies such as cDNA microarrays or proteomic approaches generate large datasets requiring biological interpretation. Text based data mining tools have recently been developed to query large biological datasets of this type of data. PubMatrix is a web-based tool that allows simple text based mining of PubMed using any two lists of keywords terms, resulting in a frequency matrix of term co-occurrence. For example, a simple term selection procedure allows automatic pairwise comparisons of approximately 1-10,000 control terms versus approximately 1-10 search terms, resulting in >100,000 pair wise comparisons. A matrix table of pair-wise comparisons can then be surveyed and queried individually. Lists of keywords can include any terms currently capable of being searched in PubMed.

In the context of cDNA microarray studies, this may be used for the annotation of gene lists from clusters of genes that are expressed coordinately. An associated PubMatrix archive provides previous searches using common useful lists of keyword terms. In this way, lists of terms, such as gene names, or functional assignments can be assigned genetic, biological or clinical relevance in a rapid, flexible systematic fashion.